CHAPTER 2: ANALYSIS OF FALL 2000 FIELD TEST DATA

Introduction

Test questions that had been developed or adapted during the first half of 2000 were included in the Spring 2000 Field Test. American Institutes for Research (AIR) reported results from that field test August of 2000 (American Institutes for Research, 2000). Our own analyses of the Spring 2000 Field Test were reported in our June 30 and August 25 reports (Wise et al., 2000a; Wise et al., 2000b). The results of the Spring 2000 Field Test indicated that nearly all of the items had acceptable statistical properties and could be used in operational CAHSEE forms. Nonetheless, additional test questions were needed to cover particular standards and to support the assembly of multiple test forms. Additional test questions were developed by AIR and included in a second field test conducted in Fall 2000. Analyses of results from this second field test are reported in this chapter.

Our analyses addressed three general issues:

- What proportion of items has good statistical properties?
- Were the questions included in the second field test significantly different in quality and difficulty compared to the questions in the first field test?
- How difficult are the questions that address specific standards and did the difficulty level vary among different demographic groups?

The answer to the first question provides an indication of the continued soundness of the development procedures and also will determine whether there are enough high-quality items to begin assembling multiple operational forms of the exam. The second wave of questions was developed from scratch specifically for the CAHSEE, while the first wave included questions selected from other sources as appropriate measures of the standards to be assessed by the CAHSEE. The newer questions had not been as extensively reviewed and we wanted to know whether this would lead to any differences in apparent quality. In addressing the third issue, we combined questions from both the Spring and Fall 2000 Field Tests.

The field test results provide an interesting contrast to the results from the March 2001 operational administration. Researchers attempted to recruit representative samples of 10th graders for the two field tests; the operational administration included 9th graders who volunteered to take the CAHSEE. With another year of schooling, 10th graders might perform significantly better on some or all parts of the tests. On the other hand, the motivation to work hard was clearly lower in the field test where students' scores would not count and would not be reported.

Field Test Design

Test Booklets

AIR constructed four test booklets (forms) of English-language arts (ELA) questions. Each form contained 120 multiple-choice (MC) questions and two constructed response (CR)

essay questions. The first essay question was included after question 82 and the second essay question was at the end of the test. A total of 43 different reading passages with MC questions (items) were tried out. Some of these passages were included in more than one test booklet with differences in some or all of the questions asked about the passage. The purpose of this repetition was to avoid asking too many different questions of any one student, but still allow the contractor to pick the best items for each passage when it is used in an operational form. In all, 372 new MC questions were tried out in the Fall Field Test, with 28 of these items included in two different forms (bringing the total MC items printed to 400, or 100 per booklet). Three different versions (subforms) of each booklet were created with the same MC questions but different essay questions. In this way a total of 24 different essay questions were tried out (2 for each of the 3 versions of each of the 4 forms).

AIR also constructed 4 forms of mathematics (math) items. Each form contained 100 MC items. There were no essay questions for math and there was no overlap across the 4 math booklets.

In order to compare results from the Fall Field Test to results from the Spring Field Test, the test developers had to adjust for differences in the achievement levels of the students who participated in the two field tests. To assess these differences, 20 ELA and 20 mathematics questions from the Spring Field Test were identified as "linking items." The 20 ELA questions were included in each of the four ELA forms. Similarly, the 20 mathematics questions were included in each of the math forms. The questions were selected to cover different standards and represent different levels of difficulty. As a result, each form in the Fall Field Test had 100 new MC questions and 20 questions from the Spring Field Test used for linking.

Field Test Sample

Details of the Field Test Sampling plan will be presented in AIR's report on the field test. The basic goal was to ensure that the sample of students completing each test booklet covered a wide range of abilities and was generally representative of 10^{th} grade students in California. Initially researchers thought that the field test would have to provide normative data for use in determining minimum passing scores. There were several limitations on this plan, including the fact that the field test participants had only just begun the 10^{th} grade curriculum and the likelihood that they may not all have been motivated to do their best. When it was believed that the 2001 administration to 9^{th} graders would be for practice only, researchers planned to collect more comprehensive normative information from a census testing of 10^{th} graders in 2002. When those plans were changed again by the failure of Senate Bill 84 (SB 84, which, in part, introduced urgency legislation that proposed delaying the requirement that students pass the CAHSEE to the class of 2005 rather than the class of 2004), data from the operational March 2001 administration were used in preference to constructing estimated passing rates from the field test results

For each of the two exams, AIR sorted California schools by their level of performance on the corresponding 2000 STAR (Standardized Testing and Reporting; for details see http://star.cde.ca.gov/) test and then picked 10 schools from the lowest performing tenth (decile) of these schools, 10 schools from the next lowest performing decile, and so on up to

10 schools from the top performing decile. This approach appears to be an effective way of obtaining samples from schools that span the full range of ELA and math abilities.

For each of the selected schools, AIR requested up to 66 tenth grade students. Some of the schools were too small to be able to comply with this request and others could not supply the total requested students because of end-of-year scheduling problems. Fortunately, AIR had been reasonably conservative in planning for this contingency and the resulting sample sizes appear adequate for most of the intended analyses. Within each school, each of the four different ELA or math booklets was assigned to roughly one-fourth of the students tested. This provided "randomly equivalent" samples of students for the different booklets (the same ability levels except for random factors in the assignment to booklet that become negligible with large sample sizes). The 20 common linking items included in each test form provided a basis for checking on the equivalence of the samples of students for each test form and also a means of equating results from each form back to the scale used in the Spring 2000 Field Test.

Table 2.1 shows the total number of students completing each booklet. In these and the tables that follow, a small number of students with missing form codes or no valid item responses were deleted from our analyses. Even though the tests were long, nearly all students responded to all of the questions. Only 5% of the ELA sample and 3% of the math sample failed to respond to (omitted) more than five of the 120 questions.

Table 2.1 also shows the average percentage of correct scores for the 120 MC questions in each form. (These numbers are also the average of the percentage of correct responses to each question.) For the ELA forms, the average score for the essay items is also shown. For both subjects, these averages and the standard deviations (which show how much the scores varied across different students) were very similar across the four test forms. Assuming that the random assignment of students to booklets worked as intended, this similarity in percent correct scores and the essay scores suggests that the questions in each of the different booklets were of comparable average difficulty.

TABLE 2.1 Average Scores by Subject and Field Test Form

Subject		Field Test Form	Sample Size	Average Score (% Correct for MC)	Standard Deviation
ELA–MC	Fall 2000	1.x	1246	56.4	19.5
		2.x	1299	58.6	20.3
		3.x	1315	56.8	19.4
		4.x	1233	56.6	19.7
		ALL	5093	57.1	19.8
	Spring 2000	ALL	3757	58.9	20.5
ELA–Essays	Fall 2000	1.x	1161	2.08*	0.61
		2.x	1213	2.11*	0.63
		3.x	1247	2.19*	0.64
		4.x	1159	2.05*	0.60
		ALL	4780	2.11*	0.62
	Spring 2000	ALL	3843	2.02	0.96
Mathematics	Fall 2000	1	1212	41.8	16.3
		2	1236	41.9	16.3
		3	1212	44.4	17.1
		4	1199	45.2	16.0
		ALL	4859	43.3	16.5
	Spring 2000	ALL	3920	47.1	18.1

* Note: Essay average reflects score out of four possible points

Item Difficulties

The results in Table 2.1 above provide important information on the average difficulty of the CAHSEE items for California students at the beginning of 10th grade. For reference, Table 2.1 also shows average rates of correct response for questions in the Spring 2000 Field Test. The Spring 2000 Field Test involved students at the very end of the 10th grade. For ELA and particularly for math, the average passing rates were somewhat lower in the Fall Field Test. This could either mean that the questions developed for the Fall Field Test were a bit more difficult or that students at the beginning of 10th grade were not as well prepared to answer the questions as were students at the end of 10th grade. See the analyses of the linking items presented below for more information on these two options.

As noted in our Spring 2000 report, all of the MC questions include four possible alternative answers. It is important to note that, because of the possibility of guessing, the percent of students who answered these questions correctly is not the same as the percent who actually knew the correct answer. For example, suppose only 25% of the students knew the correct answer to a question and the other 75% guessed randomly. All of the students knowing the answer (25%) would answer correctly and one fourth of the students who did not know the answer (18.75%) would answer correctly through random guessing, so the expected percent answering correctly would be 43.75% (25 + 18.75). Note that 43.75% is slightly **higher** than the average passing rate for the math questions (43.3%), suggesting that, on average, fewer than 25% of the students actually knew the correct answer.

We also examined the distribution of number correct scores for different demographic groups as shown in Tables 2.2 and 2.3. These results provide a preliminary indication of the relative difficulty of CAHSEE items for different groups of students. In the next chapter, we

present estimates of actual passing rates for these groups from the March 2001 administration of the tests.

TABLE 2.2 Average Total Scores by Gender

Subject	Gender	Fall 2000 N	Average Score (% Correct for MC) Fall 2000	Spring 2000
ELA–MC	Female	2535	59.5	62.9
	Male	2522	54.5	55.2
ELA-Essay	Female	2423	2.16*	2.21
	Male	2324	1.99*	1.85
Mathematics	Female	2371	43.0	46.8
	Male	2460	43.7	47.4

* Note: Essay average reflects score out of four possible points

TABLE 2.3 Average Total Scores by Race and Language Fluency

Subject	Race/Language Status	Fall 2000 N	Average Score (% Correct for MC) Fall 2000	Spring 2000
ELA–MC	African American (1)	256	51.3	50.2
	Asian (3)	490	64.2	68.8
	Hispanic (5)	1978	50.6	50.9
	White (7)	1846	64.2	65.7
	English Learners	726	42.2	40.5
ELA-Essay	African American (1)	234	1.86*	1.70
	Asian (3)	476	2.37*	2.36
	Hispanic (5)	1822	1.94*	1.80
	White (7)	1764	2.27*	2.21
	L.E.P**	638	1.78	1.45
Mathematics	African American (1)	320	35.5	41.0
	Asian (3)	345	56.4	57.6
	Hispanic (5)	1970	37.0	38.6
	White (7)	1671	49.6	52.0
	English Learners	318	35.6	14.8

* Note: Essay average reflects score out of four possible points

Table 2.4 shows the average percent of correct responses to the 20 linking items for each of the Fall Field Test Forms and for all of the students in the Spring Field Test. For ELA, the students at the beginning of 10th grade in the Fall Field Test actually did better than the students from the Spring Field Test who were at the end of the 10th grade. This suggests that the ELA questions in the Fall Field Test were, in fact, slightly easier.

The data in Table 2.4 show the opposite finding for mathematics. The sample of students at the beginning of 10th grade had lower rates of correct responses than the sample of students at the end of 10th grade by about 4.5 percentage points. The difference was about the same as the difference in passing rates for the new items, suggesting that all of the variances in percent correct were due to sample differences and that there was no difference in the average difficulty of the questions.

TABLE 2.4 Comparison of Spring and Fall Performance on Linking Items

	ELA	Mathematics
No. of Linking Items	20	20
Passing Rates in Fall 2000 Field Test	Percent Correct	Percent Correct
Form 1	63.5	52.3
Form 2	62.7	52.4
Form 3	62.4	53.3
Form 4	62.5	53.8
Fall 2000 Avg. (beginning of 10th Grade)	62.8	53.0
Spring 2000 Avg. (end of 10th Grade)	61.7	57.5
Difference	-1.1	+4.5

Item Screening

As with the Spring Field Test, we made an effort to estimate the number of field test items with statistical properties that suggest they would need to be dropped or revised (and retested) before being used in operational forms. Statistical indicators were used to assess: (a) whether items were *inappropriately* easy or difficult, (b) whether the item score provided information that was at odds with (did not generalize to) the information provided by the other items, and (c) whether the item appeared to function differently for different demographic groups (females, Hispanics, or African Americans).

Item Difficulty

We computed the percent passing (p-values) for each item. In subsequent analyses, it might be possible and desirable to adjust these p-values for differences between the field test samples and the total population of California's 10th grade students. As noted above, the procedures used in drawing the sample should have been sufficient to ensure that any such adjustments would be minor. Item difficulty screens are used to weed out items, which, although they could be perfectly valid, provide little or no useful information. More often than not, extreme item difficulties also reflect item flaws so that most of the items screened out are not valid measures of the intended standards and are inefficient as well. For example, if nearly all students pass an item, it may well be that the distracters (incorrect response options) are not plausible or that something in the item text "gives away" the correct answer. Similarly, if the percentage answering correctly is at the guessing level (suggesting that no one really knows the correct answer), the item provides little information about student skills and is likely to be flawed. In this case, the item could be incorrectly keyed or have no correct option or have some problem in the text that leads even able students astray. We flagged items with passing rates above 95% as too easy and those with passing rates below 25% (the guessing level for 4-option items) as too difficult.

Item-Total Correlation

Another indicator of potential item problems is when results from the item disagree with (fail to generalize to) the scores on other items. The item-total correlation coefficient measures the extent to which students who answer the item correctly also score well on the rest of the test. Because the item score is dichotomous (scored pass or fail) and the total score has a continuous (more normal) distribution, the range of the item-total correlations is

limited, particularly when the percentage of students passing the item is much different from 50. We computed a Clemans-Brogden biserial correlation coefficient (Lord & Novick, 1968, page 341) that corrects for differences in item difficulty. Possible values range from –1.0 to +1.0 with positive values indicating agreement between the item score and the total score. We flagged all items with values less than 0.2 as having a generalizability problem. Often these items are mis-keyed or have ambiguities in the text or options that limit their validity as a measure of achievement of the targeted standards.

Differential Item Functioning (DIF)

It is common practice to look for differences in the way an item functions across different groups of students. In most analyses of differential item functioning (DIF), a focal group is identified that is of specific concern. The rates at which members of this group answer an item correctly (pass) are compared to passing rates for a second reference group. In our analyses, Hispanics, African Americans, and females were the focal groups of interest. In each case, statistics for these students were compared to statistics for all other students in the field test.

The issue is not just whether there are different passing rates for these different groups. The question addressed in DIF analyses is whether group differences in passing rates for some items are significantly larger than the differences in passing rates for the other items. Another way of framing this issue is to ask whether students from different groups who are at the same overall level of achievement (usually indicated by the total test score) have the same probability of answering the item correctly.

We computed DIF statistics[2] for females, Hispanics and African Americans–the groups of most common concern in test bias studies. The sample sizes for females and Hispanics (more then 400 and 300 per test form respectively) were large enough to detect moderate and large DIF reliably. The sample size for African Americans was much smaller, generally 40 to 50 per item. Only a few items were flagged as having potentially significant DIF for this group, in part because the sample size was not large enough to allow detection of items with only moderate DIF.

Note that a finding of significant DIF does not necessarily mean that an item is not a valid measure of the intended standard. Group differences in preparation can lead to greater group differences on some items than on others. For example, suppose that male and female students took algebra at the same rate, but many more male students went on to take geometry by the 10th grade. We would expect larger gender differences in passing rates for

[2] A commonly used DIF statistic, the Mantel-Haenszel log odds ratio (Mantel & Haenszel, 1959), compares the odds of passing the item (percent correct/percent incorrect) for focal and reference group members at each different total score level. An odds ratio is computed for each total score level (indicating comparable overall ability). If the odds of passing for the focal group are the same as for the reference group, the ratio of the odds values is 1.0 and the logarithm of this ratio is 0.0. To the extent that the log odds values (across all of the score levels) are different from 0.0, the item is said to function differently (be disproportionately hard or easy) for the focal and reference groups. We computed a chi-square statistic (see Dorans & Holland, 1993, page 40) that tests whether the Mantel-Haenszel statistic is different from 0.0. We flagged cases where the statistic was greater than 7.8794. This corresponds to the .005 level for a one-degree chi-square, meaning that there was less than .01 chance of getting a value this large (or a correspondingly small one) by chance alone.

geometry items than for algebra items, even if all items were perfectly valid measures of their intended content. A common practice is to flag all items with significant DIF values for further content and sensitivity review. Many of these items would then be accepted and used without further changes. We used a relatively high cutoff (the .01 level of statistical significance) to estimate the proportion of items that would eventually be screened out because of DIF concerns.

Item Screening Results

Table 2.5 summarizes our item screening results. It should be noted that these are preliminary estimates based on statistical criteria only. AIR will end up with somewhat different results using somewhat different statistical criteria and incorporating editorial, as well as statistical, review of flagged items.

Overall the results show the items developed for the Fall 2000 Field Test were of good quality with relatively few questions flagged for statistical concerns. In many programs, half of the items or more are screened out on the basis of initial field test results. We flagged only 1 out of 4 of the math items and 1 out of 8 of the ELA items. The high survival rates for the questions in the Fall Field Test were only slightly lower than for the Spring Field Test, even though most of the items in the Spring Field Test had been previously screened. The high survival rates indicate a high level of effectiveness in the item development and review procedures.

TABLE 2.5 Percent of Multiple-Choice (MC) Items Screened Out by Various Statistical Criteria

Subject/Statistic	ELA-MC	Math
Total new field test items Fall 2000	352	400
Number passing all screens Fall 2000	297	288
Percent passing all screens Fall 2000	84.4%	72.0%
% Too easy*	0.0%	0.0%
% Too hard	1.7%	11.8%
% Low item-total correlation	3.1%	13.0%
% DIF–Female	7.4%	7.0%
% DIF–Hispanic	6.0%	0.8%
% DIF–African American	0.3%	0.3%
Percent passing all screens Spring 2000	87.0%	77.5%

* Note: Percents add to more than 100 because some items were flagged for more than one reason

Relative Difficulty of Questions by Content Standard

In our August 2000 report, we looked at the relative difficulty of the questions developed to assess each content standard. With the completion of the Fall Field Test, there are now twice as many questions for each content standard, providing a more extensive basis for comparing the relative difficulty of the different standards. Tables 2.6 through 2.8 show the number of items developed for each content standard and the average percent passing for these items. The number of questions to be included in each test form is also shown, providing a basis for determining the number of different test forms that might be assembled from the items "surviving" the Spring and Fall Field Tests. Tables 2.6, 2.7 and 2.8 reference California Content Standards in the following format:

1.2 Distinguish between the denotative and connotative meanings of words and interpret the connotative power of words.

These standards, including the standard number (1.2 in the example), can be found in the *CAHSEE Language Arts Blueprint* (see http://www.cde.ca.gov/ta/tg/hs/) and the *CAHSEE Mathematics Blueprint (see* http://www.cde.ca.gov/ta/tg/hs/), both of which were approved by the State Board of Education on December 7, 2000. These standards comprise a subset of the complete set of CDE standards. Missing standards (as evidenced by gaps in the sequence) are not included in the CAHSEE exam and have been omitted here deliberately. Some standards (e.g., Table 2.8 Standard 24.3) were included in the field test but subsequently eliminated from CAHSEE; their statistics were included in the tables although the "number per form" is zero.

TABLE 2.6 Number and Difficulty of CAHSEE Questions by Test Content Standard: Language Arts—Reading

STRANDS/STANDARDS [BOLDED STANDARDS INDICATE PASS RATES of 50% OR LOWER]	No. per Form	No. in Bank	Average % Pass
1.0 Word Analysis, Fluency, and Systematic Vocabulary Development (Grades 9-10)			
Vocabulary and Concept Development			
1.1 Identify and use the literal and figurative meanings of words and understand word derivations.	6	64	62%
1.2 Distinguish between the denotative and connotative meanings of words and interpret the connotative power of words.	4	8	63%
2.0 Reading Comprehension (Focus on Informational Materials) (Grade 9-10 except as noted)			
Structural Features of Informational Materials			
8.2.1 Compare and contrast the features and elements of consumer materials to gain meaning from documents (e.g., warranties, contracts, product information, instructional manuals). [NOTE: This is a grade eight standard.]	1	8	66%
2.1 Analyze the structure and format of functional workplace documents, including the graphics and headers, and explain how authors use the features to achieve their purposes.	3	22	73%
2.2 Prepare a bibliography of reference materials for a report using a variety of consumer, workplace, and public documents.	2	5	49%
Comprehension and Analysis of Grade-Level-Appropriate Text			
2.3 Generate relevant questions about readings on issues that can be researched.	2	12	**50%**
2.4 Synthesize the content from several sources or works by a single author dealing with a single issue; paraphrase the ideas and connect them to other sources and related topics to demonstrate	3	77	63%

TABLE 2.6 Number and Difficulty of CAHSEE Questions by Test Content Standard: Language Arts—Reading

STRANDS/STANDARDS [BOLDED STANDARDS INDICATE PASS RATES of 50% OR LOWER]	No. per Form	No. in Bank	Average % Pass
comprehension.			
2.5 Extend ideas presented in primary or secondary sources through original analysis, evaluation, and elaboration.	3	47	59%
2.7 Critique the logic of functional documents by examining the sequence of information and procedures in anticipation of possible reader misunderstandings.	3	14	63%
2.8 Evaluate the credibility of an author's argument or defense of a claim by critiquing the relationship between generalizations and evidence, the comprehensiveness of evidence, and the way in which the author's intent affects the structure and tone of the text (e.g., in professional journals, editorials, political speeches, primary source material).	7	33	56%
3.0 Literary Response and Analysis (Grades 9-10):			
Structural Features of Literature			
3.1 Articulate the relationship between the expressed purposes and the characteristics of different forms of dramatic literature (e.g., comedy, tragedy, drama, dramatic monologue).	2	8	65%
Narrative Analysis of Grade-Level-Appropriate Text			
3.3 Analyze interactions between main and subordinate characters in a literary text (e.g., internal and external conflicts, motivations, relationships, influences) and explain the way those interactions affect the plot.	2	16	66%
3.4 Determine characters' traits by what the characters say about themselves in narration, dialogue, dramatic monologue, and soliloquy.	2	25	61%
3.5 Compare works that express a universal theme and provide evidence to support the ideas expressed in each work.	4	13	61%
3.6 Analyze and trace an author's development of time and sequence, including the use of complex literary devices (e.g., foreshadowing, flashbacks).	2	7	53%
3.7 Recognize and understand the significance of various literary devices, including figurative language, imagery, allegory, and symbolism, and explain their appeal.	2	13	51%
3.8 Interpret and evaluate the impact of ambiguities, subtleties, contradictions, ironies, and incongruities in a text.	2	10	59%
3.9 Explain how voice, persona, and the choice of a narrator affect characterization and the tone, plot, and credibility of a text.	2	6	**50%**
3.10 Identify and describe the function of dialogue, scene designs, soliloquies, asides, and character foils in dramatic literature.	2	9	52%

TABLE 2.6 Number and Difficulty of CAHSEE Questions by Test Content Standard: Language Arts—Reading

STRANDS/STANDARDS [BOLDED STANDARDS INDICATE PASS RATES of 50% OR LOWER]	No. per Form	No. in Bank	Average % Pass
Literary Criticism			
8.3.7 Analyze a work of literature, showing how it reflects the heritage, traditions, attitudes, and beliefs of its author. (Biographical approach) [NOTE: This is a grade eight standard.]	1.3*	4	71%
3.11 Evaluate the aesthetic qualities of style, including the impact of diction and figurative language on tone, mood, and theme, using the terminology of literary criticism. (Aesthetic approach)	1.3*	3	71%
3.12 Analyze the way in which a work of literature is related to the themes and issues of its historical period. (Historical approach)	1.3*	1	54%

- Note: 4 questions rotated across approaches in different test forms

TABLE 2.7 Number and Difficulty of CAHSEE Questions by Test Content Standard: Language Arts—Writing

STRANDS/STANDARDS [BOLDED STANDARDS INDICATE PASS RATES of 50% OR LOWER]	No. per Form	No. in Bank	Average % Pass
1.0 Writing Strategies (Grades 9-10):			
Organization and Focus			
1.1 Establish a controlling impression or coherent thesis that conveys a clear and distinctive perspective on the subject and maintain a consistent tone and focus throughout the piece of writing.	2	23	54%
1.2 Use precise language, action verbs, sensory details, appropriate modifiers, and the active rather than the passive voice.	3	17	54%
Research and Technology			
1.3 Use clear research questions and suitable research methods (e.g., library, electronic media, personal interview) to elicit and present evidence from primary and secondary sources.	1	9	62%
1.4 Develop the main ideas within the body of the composition through supporting evidence (e.g., scenarios, commonly held beliefs, hypotheses, definitions).	1	11	53%
1.5 Synthesize information from multiple sources and identify complexities and discrepancies in the information and the different perspectives found in each medium (e.g., almanacs, microfiche, news sources, in-depth field studies, speeches, journals, technical documents).	1	2	55%
1.6 Integrate quotations and citations into a written text while maintaining the flow of ideas.	1	5	54%
1.9 Revise writing to improve the logic and coherence of the organization and controlling perspective, the precision of word choice, and the tone by taking into consideration the audience, purpose, and formality of the context.	2	4	52%

TABLE 2.7 Number and Difficulty of CAHSEE Questions by Test Content Standard: Language Arts—Writing

STRANDS/STANDARDS [BOLDED STANDARDS INDICATE PASS RATES of 50% OR LOWER]	No. per Form	No. in Bank	Avg. Score (range 0-4)
2.0 Writing Applications (Genres and Their Characteristics)			
2.1 Write biographical narratives: a. Relate a sequence of events and communicate the significance of the events to the audience. b. Locate scenes and incidents in specific places. c. Describe with concrete sensory details the sights, sounds, and smells of a scene and the specific actions, movements, gestures, and feelings of the characters; use interior monologue to depict the characters' feelings. e. Make effective use of descriptions of appearance, images, shifting perspectives, and sensory details.	.33** essay	5	2.05
2.2. Write responses to literature: a. Demonstrate a comprehensive grasp of the significant ideas of literary works. b. Support important ideas and viewpoints through accurate and detailed references to the text or to other works. c. Demonstrate awareness of the author's use of stylistic devices and an appreciation of the effects created. d. Identify and assess the impact of perceived ambiguities, nuances and complexities within the text.	0.5*** essay	12	2.11
2.3 Write expository compositions, including analytical essays and research reports: a. Marshal evidence in support of a thesis and related claims, including information on all relevant perspectives. b. Convey information and ideas from primary and secondary sources accurately and coherently. c. Make distinctions between the relative value and significance of specific data, facts, and ideas. e. Anticipate and address readers' potential misunderstandings, biases, and expectations. f. Use technical terms and notations accurately.	0.5*** essay	7	2.07
2.4 Write persuasive compositions: a. Structure ideas and arguments in a sustained and logical fashion. b. Use specific rhetorical devices to support assertions (e.g., appeal to logic through reasoning; appeal to emotion or ethical belief; relate a personal anecdote, case study, or analogy). c. Clarify and defend positions with precise and relevant evidence, including facts, expert opinions, quotations, and expressions of commonly accepted beliefs and logical reasoning. d. Address readers' concerns, counterclaims, biases, and expectations.	0.33** essay	12	1.98

TABLE 2.7 Number and Difficulty of CAHSEE Questions by Test Content Standard: Language Arts—Writing

STRANDS/STANDARDS [BOLDED STANDARDS INDICATE PASS RATES of 50% OR LOWER]	No. per Form	No. in Bank	Avg. Score (range 0-4)
2.5 Write business letters: a. Provide clear and purposeful information and address the intended audience appropriately. b. Use appropriate vocabulary, tone, and style to take into account the nature of the relationship with, and the knowledge and interests of, the recipients. c. Highlight central ideas or images. d. Follow a conventional style with page formats, fonts, and spacing that contribute to the document's readability and impact.	0.33** essay	12	2.05

STRANDS/STANDARDS [BOLDED STANDARDS INDICATE PASS RATES of 50% OR LOWER]	No. per Form	No. in Bank	Average % Pass
1.0 Written and Oral English Language Conventions (Grades 9 & 10):			
Grammar and Mechanics of Writing			
1.1 Identify and correctly use clauses (e.g., main and subordinate), phrases (e.g., gerund, infinitive, and participial), and mechanics of punctuation (e.g., semicolons, colons, ellipses, hyphens).	4	27	54%
1.2 Understand sentence construction (e.g., parallel structure, subordination, proper placement of modifiers) and proper English usage (e.g., consistency of verb tenses).	4	34	58%
1.3 Demonstrate an understanding of proper English usage and control of grammar, paragraph and sentence structure, diction, and syntax.	4	52	60%
Manuscript Form			
1.5 Reflect appropriate manuscript requirements, including title page presentation, pagination, spacing and margins, and integration of source and support material (e.g., in-text citation, use of direct quotations, paraphrasing) with appropriate citations.	1	9	**48%**

- ** Note: 3 questions rotated across different test forms
- ***Note: 2 questions rotated across different test forms

TABLE 2.8 Number and Difficulty of CAHSEE Questions by Test Content Standard: Mathematics

STRANDS/STANDARDS [BOLDED STANDARDS INDICATE PASS RATES of 50% OR LOWER]	No. per Form	No. in Bank	Average % Pass
Grade 6—Statistics, Data Analysis, and Probability			
1.0 Students compute and analyze statistical measurements for data sets:			
1.1 Compute the mean, median, and mode of data sets.	1	15	**50%**

TABLE 2.8 Number and Difficulty of CAHSEE Questions by Test Content Standard: Mathematics

STRANDS/STANDARDS [BOLDED STANDARDS INDICATE PASS RATES of 50% OR LOWER]	No. per Form	No. in Bank	Average % Pass
2.0 Students use data samples of a population and describe the characteristics and limitations of the samples:			
2.5 Identify claims based on statistical data and, in simple cases, evaluate the validity of the claims.	1	6	52%
3.0 Students determine theoretical and experimental probabilities and use these to make predictions about events:			
3.1 Represent all possible outcomes for compound events in an organized way (e.g., tables, grids, tree diagrams) and express the theoretical probability of each outcome.	1	7	**44%**
3.3 Represent probabilities as ratios, proportions, decimals between 0 and 1, and percentages between 0 and 100, and verify that the probabilities computed are reasonable; know that if P is the probability of an event, 1-P is the probability of an event not occurring.	2	15	54%
3.5 Understand the difference between independent and dependent events.	1	7	**41%**
Grade 7—Number Sense			
1.0 Students know the properties of, and compute with, rational numbers expressed in a variety of forms:			
1.1 Read, write, and compare rational numbers in scientific notation (positive and negative powers of 10) with approximate numbers using scientific notation.	1	5	54%
1.2 Add, subtract, multiply, and divide rational numbers (integers, fractions, and terminating decimals) and take positive rational numbers to whole-number powers.	3	19	61%
1.3 Convert fractions to decimals and percents and use these representations in estimations, computations, and applications.	2	7	**50%**
1.6 Calculate the percentage of increases and decreases of a quantity.	1	3	**43%**
1.7 Solve problems that involve discounts, markups, commissions, and profit, and compute simple and compound interest.	2	9	51%
2.0 Students use exponents, powers, and roots, and use exponents in working with fractions:			
2.1 Understand negative whole-number exponents. Multiply and divide expressions involving exponents with a common base.	1	4	**33%**
2.2 Add and subtract fractions by using factoring to find common denominators.	1	7	**39%**
2.3 Multiply, divide, and simplify rational numbers by using exponent rules.	1	8	58%

TABLE 2.8 Number and Difficulty of CAHSEE Questions by Test Content Standard: Mathematics

STRANDS/STANDARDS [BOLDED STANDARDS INDICATE PASS RATES of 50% OR LOWER]	No. per Form	No. in Bank	Average % Pass
2.4 Use the inverse relationship between raising to a power and extracting the root of a perfect square integer; for an integer that is not square, determine without a calculator the two integers between which its square root lies and explain why.	1	10	**49%**
2.5 Understand the meaning of the absolute of a number; interpret the absolute value as the distance of the number from zero on a number line; and determine the absolute value of real numbers.	1	7	64%
Grade 7—Algebra and Functions			
1.0 Students express quantitative relationships by using algebraic terminology, expressions, equations, inequalities, and graphs:			
1.1 Use variables and appropriate operations to write an expression, an equation, an inequality, or a system of equations or inequalities that represents a verbal description (e.g., three less than a number, half as large as area A).	2	9	55%
1.2 Use the correct order of operations to evaluate [simplify] algebraic expressions such as $3\ (2x+5)^{2.}$	1	5	55%
1.5 Represent quantitative relationships graphically and interpret the meaning of a specific part of a graph in the situation represented by the graph.	3	14	59%
2.0 Students interpret and evaluate expressions involving integer powers and simple roots:			
2.1 Interpret positive whole-number powers as repeated multiplication and negative whole-number powers as repeated division or multiplication by the multiplicative inverse. Simplify and evaluate expressions that include exponents.	1	5	53%
2.2 Multiply and divide monomials; extend the process of taking powers and extracting roots to monomials when the latter results in a monomial with an integer exponent.	1	4	**44%**
3.0 Students graph and interpret linear and some nonlinear functions:			
3.1 Graph functions of the form $y=nx^2$ and $y=nx^3$ and use in solving problems.	1	6	**38%**
3.3 Graph linear functions, noting that the vertical change (change in y-value) per unit of horizontal change (change in x -value) is always the same and know that the ratio ("rise over run") is called the slope of a graph.	2	13	**45%**
3.4 Plot the values of quantities whose ratios are always the same (e.g., cost to the number of an item, feet to inches, circumference to diameter of a circle). Fit a line to the plot and understand that the slope of a line equals the [ratio of the] quantities.	1	6	55%

TABLE 2.8 Number and Difficulty of CAHSEE Questions by Test Content Standard: Mathematics

STRANDS/STANDARDS [BOLDED STANDARDS INDICATE PASS RATES of 50% OR LOWER]	No. per Form	No. in Bank	Average % Pass
4.0 Students solve simple linear equations and inequalities over the rational numbers:			
4.1 Solve two-step linear equations and inequalities in one variable over the rational numbers, interpret the solution or solutions in the context from which they arose, and verify the reasonableness of the results.	3	20	59%
4.2 Solve multistep problems involving rate, average speed, distance, and time, or a direct variation.	2	11	56%
Grade 7—Measurement and Geometry			
1.0 Students choose appropriate units of measure and use ratios to convert within and between measurement systems to solve problems:			
1.1 Compare weights, capacities, geometric measures, times, and temperatures within and between measurement systems (e.g., miles per hour and feet per second, cubic inches to cubic centimeters).	2	18	**50%**
1.2 Construct and read drawings and models made to scale.	1	8	**41%**
1.3 Use measures expressed as rates (e.g., speed, density) and measures expressed as products (e.g., person-days) to solve problems; check the units of the solutions; and use dimensional analysis to check the reasonableness of the answer.	2	12	61%
2.0 Students compute the perimeter, area, and volume of common geometric objects and use the results to find measures of less common objects. They know how perimeter, area and volume are affected by changes of scale:			
2.1 Use formulas routinely for finding the perimeter and area of basic two-dimensional figures and the surface area and volume of basic three-dimensional figures, including rectangles, parallelograms, trapezoids, squares, triangles, circles, prisms, and cylinders.	3	22	**43%**
2.2 Estimate and compute the [surface] area of more complex or irregular two- and three-dimensional figures by breaking the figures down into more basic geometric objects.	2	13	**44%**
2.3 Compute the length of the perimeter, the surface area of the faces, and the volume of a three-dimensional object built from rectangular solids. Understand that when the lengths of all dimensions are multiplied by a scale factor, the surface area is multiplied by the square of the scale factor and the volume is multiplied by the cube of the scale factor.	1	7	**44%**
2.4 Relate the changes in measurement with a change of scale to the units used (e.g., square inches, cubic feet) and to conversions between units (1 square foot = 144 square inches or [1 ft^2] = {144 in^2}, 1 cubic inch is approximately 16.38 cubic centimeters or [1 in^3] = [16.38 cm^3].	1	5	51%

TABLE 2.8 Number and Difficulty of CAHSEE Questions by Test Content Standard: Mathematics

STRANDS/STANDARDS [BOLDED STANDARDS INDICATE PASS RATES of 50% OR LOWER]	No. per Form	No. in Bank	Average % Pass
3.0 Students know the Pythagorean theorem and deepen their understanding of plane and solid geometric shapes by constructing figures that meet given conditions and by identifying attributes of figures:			
3.2 Understand and use coordinate graphs to plot simple figures, determine lengths and areas related to them, and determine their images under translations and reflections.	2	13	**44%**
3.3 Know and understand the Pythagorean theorem and its converse and use it to find the length of the missing side of a right triangle and the lengths of other line segments and, in some situations, empirically verify the Pythagorean theorem by direct measurement.	2	15	**37%**
3.4 Demonstrate an understanding of conditions that indicate two geometrical figures are congruent and what congruence means about relationships between the sides and angles of the two figures.	1	5	52%
Grade 7—Statistics, Data Analysis, and Probability			
1.0 Students collect, organize, and represent data sets that have one or more variables and identify relationships among variables within a data set by hand and through the use of an electronic spreadsheet software program:			
1.1 Know various forms of display for data sets, including a stem-and-leaf plot or box-and-whisker plot; use the forms to display a single set of data or to compare two sets of data.	2	14	58%
1.2 Represent two numerical variables on a scatter plot and informally describe how the data points are distributed and any apparent relationship that exists between the two variables (e.g., between time spent on homework and grade level).	2	17	58%
1.3 Understand the meaning of, and be able to compute the minimum, the lower quartile, the median, the upper quartile, and the maximum of a data set.	2	19	**49%**
Grade 7—Mathematical Reasoning			
1.0 Students make decisions about how to approach problems:			
1.1 Analyze problems by identifying relationships, distinguishing relevant from irrelevant information, identifying missing information, sequencing and prioritizing information, and observing patterns.	2	14	58%
1.2 Formulate and justify mathematical conjectures based on a general description of the mathematical question or problem posed.	1	6	**40%**
2.0 Students use strategies, skills, and concepts in finding solutions:			
2.1 Use estimation to verify the reasonableness of calculated results.	1	6	58%

TABLE 2.8 Number and Difficulty of CAHSEE Questions by Test Content Standard: Mathematics

STRANDS/STANDARDS [BOLDED STANDARDS INDICATE PASS RATES of 50% OR LOWER]	No. per Form	No. in Bank	Average % Pass
2.3 Estimate unknown quantities graphically and solve for them by using logical reasoning and arithmetic and algebraic techniques.	1	3	**49%**
2.4 Make and test conjectures by using both inductive and deductive reasoning.	1	14	**49%**
3.0 Students determine a solution is complete and move beyond a particular problems by generalizing to other situations:			
3.1 Evaluate the reasonableness of the solution in the context of the original.	1	8	55%
3.3 Develop generalizations of the results obtained and the strategies used and apply them to new problem situations.	1	7	55%
Algebra 1			
2.0 Students understand and use such operations as taking the opposite, finding the reciprocal, and taking a root. They understand and use the rules of exponents.	1	13	**44%**
3.0 Students solve equations and inequalities involving absolute values.	1	6	**34%**
4.0 Students simplify expressions before solving linear equations and inequalities in one variable, such as 3(2x-5) + 4(x-2) = 12.	2	10	**44%**
5.0 Students solve multistep problems, including word problems, involving linear equations and linear inequalities in one variable and provide justification for each step.	1	22	**42%**
6.0 Students graph a linear equation and compute the x- and y-intercepts (e.g., graph 2x + 6y = 4).	2	15	**36%**
7.0 Students verify that a point lies on a line, given an equation of the line. Students are able to derive linear equations.	1	16	**36%**
8.0 Students understand the concepts of parallel lines and how those slopes are related.	1	10	**38%**
9.0 Students solve a system of two linear equations in two variables algebraically and are able to interpret the answer graphically. Students are able to solve a system of two linear inequalities in two variables and to sketch the solution sets.	1	6	**43%**
10.0 Students add, subtract, multiply, and divide monomials and polynomials. Students solve multi-step problems, including word problems, by using these techniques.	1	5	**36%**
15.0 Students apply algebraic techniques to solve rate problems, work problems and percent-mixture problems.	1	8	**38%**
16.0 Students understand the concepts of a relation and a function, determine whether a given relation defines a function, and give pertinent information about given relations and functions.	0	2	**28%**

TABLE 2.8 Number and Difficulty of CAHSEE Questions by Test Content Standard: Mathematics

STRANDS/STANDARDS [BOLDED STANDARDS INDICATE PASS RATES of 50% OR LOWER]	No. per Form	No. in Bank	Average % Pass
17.0 Students determine the domain of independent variables and the range of dependent variables defined by a graph, a set of ordered pairs, or a symbolic expression [an equation].	0	3	**36%**
18.0 Students determine whether a relation defined by a graph, a set of ordered pairs, or a symbolic expression [an equation] is a function and justify the conclusion.	0	2	**33%**
21.0 Students graph quadratic functions and know that their roots are the x-intercepts.	0	2	**38%**
23.0 Students apply quadratic equations to physical problems, such as the motion of an object under the force of gravity.	0	7	**36%**
24.0 Students use and know simple aspects of a logical argument:			
24.2 Students identify the hypothesis and conclusion in logical deduction.	0	3	**40%**
24.3 Students use counterexamples to show that an assertion is false and recognize that a single counterexample is sufficient to refute an assertion.	0	2	**27%**
25.0 Students use properties of the number system to judge the validity of results, to justify each step of a procedure, and to prove or disprove statements: **25.1 Students use properties of numbers to construct simple, valid arguments (direct and indirect) for, or formulate counterexamples to, claimed assertions.** **25.2 Students judge the validity of an argument according to whether the properties of the real number system and the order of operations have been applied correctly at each step.** **25.3 Given a specific algebraic statement involving linear, quadratic, or absolute value expressions or equations or inequalities, students determine whether the statement is true sometimes, always, or never.**	0	14	**37%**

Summary

Overall, the results from the CAHSEE field test were quite positive. Notwithstanding the fact that the test was quite long, nearly all the students answered all of the items. The sample sizes, while smaller than hoped for, were adequate to provide stable estimates of both traditional and IRT item parameters. One limitation was the relatively modest number of African Americans, students with disabilities, and English learner (EL) students who were tested, making it difficult to determine whether the items functioned differently for these groups.

Relatively few items had obvious statistical problems. These results were similar to the results from the Spring Field Test and confirmed results of direct observation that the process of item development and review was thorough and effective.

Comparison between percent correct statistics for the Spring and Fall Field Tests suggests that, for mathematics at least, students are increasing their level of achievement during the 10th grade. The lack of difference for ELA is puzzling. Of course, neither sample was necessarily representative of 10th graders as a whole. Much more definitive information should be available from the test results in 2002, when students who did not pass the exams as 9th graders are retested during the 10th grade.

One concern raised by the field test results was the relative difficulty of the items, particularly in mathematics. If these items reflect what we believe students need to know and be able to do, and several panels of reviewers believe that they do, then a significant number of 10th grade students are likely to fail this exam. Comparative pass rates of correct responses shown above indicate that groups who traditionally score lower on assessments of student achievement will fail at higher rates. It will be important, therefore, to ensure that there are effective programs to help students at risk, both before and after their initial experience with this exam. It is possible that students will perform at higher levels during operational testing than they did on this field test where the results do not count. However, the very high completion rates suggest that nearly all students took the field test seriously.